NSW Treasury

Sampling Strategy

TPG 22-22 Policy and Guidelines: Evaluation – Technical Note

June 2023

Acknowledgement of Country

We acknowledge that Aboriginal and Torres Strait Islander peoples are the First Peoples and Traditional Custodians of Australia, and the oldest continuing culture in human history. We pay respect to Elders past and present and commit to respecting the lands we walk on, and the communities we walk with.

Artwork:
Regeneration by Josie Rose

Contents

Abstract

Technical Note: Sampling Strategy	
Background	A sampling strategy is required to gather evidence about a target population. Sampling is the selection of a subset of individuals or units (the sample) from within a population, to draw conclusions about the whole population.
Scope	There are two main approaches to sampling, probability sampling and non-probability sampling. Key considerations in developing the sampling approach include: • aim of the research • feasibility of sampling • sample composition • sample size • rate of response.
When to use this technical note?	When trying to draw conclusions about a population through a sample, or understanding deeper insights in relation to a research question about a specific group.
Potential implications	Appropriate sample size and sample composition can reduce sample error and sample bias.
Keywords	Population, Sampling frame, Sample, Respondent, Probability sampling, Non-probability sampling, Weighting, Bias
Associated resources	For further information on sampling approaches and how/where to apply them, see the Australian Bureau of Statistics, Sample Design.

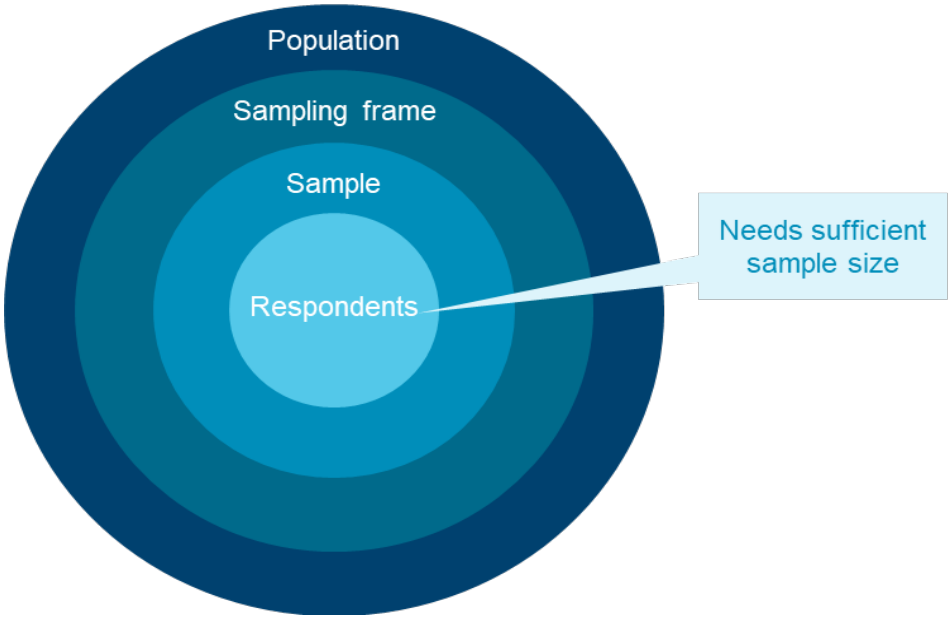
Context

Obtaining information from every member of a target population is often not feasible nor necessary. To collect data from a sample, researchers first need to identify who in the population might be realistically contactable. Data may then be captured from whomever in the sample responds or participates. Key terms are outlined in Table 1 and illustrated in Figure 1.

Table 1: Explanation of terms

Term	Description	Example
Population	The entire group that the study aims to draw conclusions about, which includes the intervention group and the control/comparison group.	All primary school students in New South Wales.
Sampling frame	The list of units from which data can be collected. The sampling frame would ideally be the same as the population; however, in some cases, data may be missing, imperfect or not up to date and thus the sampling frame may not capture all units of the population.	All primary school students for which a contact number is listed and up to date.
Sample	A subgroup of the population drawn from the sampling frame, from which quantitative or qualitative data are sought.	All primary school students from the sampling frame that are from schools A, B and C only.
Respondents (sample)	A subgroup of the sample, from which quantitative or qualitative data are captured. This group is often also referred to as the sample. If there is a 100 per cent response rate, then the group of respondents are equivalent to the full sample from which data are sought.	All primary school students from the sample who were given parental permission to participate in the survey.

Figure 1: Illustrative demonstration of population, sampling frame, sample and respondents



Note: There may not be differences between the population, sampling frame, sample, and respondents, depending on the research. For instance, the sampling frame may contain the full population where all contact details are available; the sample may contain the full sampling frame, where data is sought from the full sampling frame; and the respondents may contain the full sample, where there is a 100 per cent response rate.

Types of Sampling

Probability sampling

Probability sampling randomly selects the sample, so that it is more likely to be an unbiased representation of the population. The aim of probability sampling is to generalise findings to the broader population. Probability sampling approaches are provided in Table 2 and illustrated in Figure 2[1].

Table 2: Probability sampling approaches

Approach	Description	Example
Simple random sampling	Randomly choosing a sample among a population	Selecting all students from just two schools in New South Wales.
Cluster sampling	Randomly choosing a sample among clusters of the population, where the units are grouped by some characteristic	Selecting one school from each local government area (LGA) in New South Wales.
Stratified sampling	Randomly choosing a sample from smaller groups (strata) within a population.	Selecting a number of females from one school in each LGA in New South Wales, based on the total population of that LGA relative to the whole state.

Figure 2: Probability sampling approaches

Non-probability sampling

Non-probability sampling uses non-random approaches to select a sample. The researcher selects the sample based on a set of criteria. The aim of non-probability sampling may vary depending on the research. It should not be used to generalise findings about a population.

These techniques are not likely to produce samples that are statistically representative of the population. They may allow for deeper insights into a research question about a specific group. They may be appropriate where the sampling frame is small (which restricts the benefits and feasibility of probability sampling).

Examples of non-probability sampling approaches:

- Purposive sampling: selecting participants based on the purpose of the study
- Snowball sampling: asking participants to suggest other participants
- Convenience sampling: selecting participants based on their availability
- Self-selection sampling: inviting participants to self-select into the study.

[1] https://www.scribbr.com/methodology/sampling-methods/ (accessed 19 May 2023)

Methods

Key considerations

Aim of the research

The aim of the research should influence the sampling strategy:

- Where the aim of the research is to generalise findings from the sample to the target population, probability sampling may be more appropriate.
- Where the aim of the research is to investigate a certain cohort's outcomes in depth, non-probability sampling may be more appropriate. Where non-probability sampling is undertaken, the aim of the research may also influence the preferred sample composition and size (see sections below).

Feasibility

The sampling approach, including whether probability or non-probability sampling is used, may be limited by feasibility. Evaluators should consider:

- The costs and timeframes associated with data collection.
- The size of the sampling frame (for instance, a small or unknown sampling frame reduces the chances of selecting a random sample).
- The expected response rate (for instance, only 100 responses may be received from a survey sent to a sample of 3,000).
- The diminishing marginal benefits to increasing the sample size (for instance, an increase in sample size from 30 to 60 may have a significant impact on results; while an increase in sample size from 330 to 360 may have a minimal impact).

Sample composition

The sample composition is the structure of the sample, or what the sample looks like in terms of key characteristics (e.g., age, gender, sector). A 'representative sample' means the sample is similar to the population in key characteristics relevant to the research.

In probability sampling, the composition of the sample is likely to represent the composition of the population, as units in the sample are chosen randomly.

In non-probability sampling, the composition of the sample may or may not be representative of the population, depending on the aim of the research. For instance, if the aim of the research is to understand outcomes in a specific cohort, the sample composition may be constructed to contain a larger proportion of participants from this cohort.

Sample size

The sample size is the number of units (e.g., individuals, households or businesses) included in a sample.

The preferred sample size depends on a range of factors:

- In probability sampling, the sample size needs to be large enough to draw statistically significant conclusions and to ensure validity of findings.
- In non-probability sampling, the sample size will depend on the aim of the evaluation and the context of the initiative.

Probability sampling

Online resources can be used to determine the required sample size for a population. The Australian Bureau of Statistics Sample Size Calculator is useful for random samples.

Table 3 outlines the key factors that influence the required sample size, and the standard values that may be used in the calculators.

Table 3: Factors that influence the required sample size (probability sampling)

Term	Description	Sampling considerations
Statistical power	Also known as 'the power of a hypothesis test'; measures the probability that a researcher will find an effect in a sample if the effect exists in the population.[2] Statistical power will increase as the sample size increases.	It is good practice to set the statistical power to no less than 80 per cent. If the effect size of an intervention is expected to be large, it is possible to achieve 80 per cent power with a smaller sample size. However, if the effect size is expected to be small, a larger sample may be needed.
Effect size (scale of change)	The scale of change attributed to the initiative. Experimental and quasi-experimental evaluation designs can identify the effect size through comparing the difference in outcomes between the intervention and control or comparison groups after the initiative has been implemented. One method involves calculating the difference in means (averages) between the two groups, divided by the standard deviation (a measure of the dispersion of values) of the sampling frame.	
Confidence level	The probability that a result in the sample would represent the result in the population.	It is good practice to set the confidence level to 95 per cent. This level means that if a researcher sampled the population 20 times, they would expect the result to sit inside the confidence interval in 19 out of 20 samples. In general, higher confidence levels mean wider confidence intervals. The wider the interval, the greater the likelihood the sample estimate includes the true population value. Researchers will need to consider what confidence interval suits their problem, noting that larger samples are needed to increase the confidence level or lower the confidence interval.
Confidence interval/ Margin of error	A range of values that provide the upper and lower bounds of the result in the population, for a given confidence level. The larger the range, the less certainty there is about the result in the population.	
Variability in outcomes	The expected range (variability) in results in the population. Where data are available, standard deviation can be used to measure the level of variation.	Where values are expected to vary widely, a larger sample will assist in identifying what is the 'usual variation' of the outcome and what is the change caused by the initiative.

[2] Power calculations are based on parameters including the effect size, variability in outcomes, the confidence level and the sample size. If the sample size is pre-determined, the evaluator can use a formula to calculate the statistical power associated with the sample. Otherwise, the statistical power should be set to 80 per cent to calculate the sample size needed.

Term	Description	Sampling considerations
Population size	The number of people in the entire group that the study aims to draw conclusions about.	If a population is large, a relatively large sample size is needed. If a population is small, a smaller sample is needed (although this smaller sample may make up a larger proportion of the population).

Where the number of respondents is not sufficiently large (to achieve the required statistical power, confidence level and confidence interval), limitations should be noted in the evaluation report.

Non-probability sampling

Non-probability sampling is predominantly used for qualitative descriptions, based on interviews and survey responses. Table 4 outlines key factors that may inform the required sample size.

Table 4: Factors that influence the required sample size (non-probability sampling)

Term	Description	Sampling considerations	Example
Aim of the research	The reason for the evaluation, or the evaluation questions that need to be answered.	If the evaluation questions require detailed responses, consider focusing the research on a smaller number of stakeholders who can provide this information.	If the evaluation questions focus on a specific topic such as risk management processes, a smaller sample size may be needed to obtain the necessary information to answer the evaluation question.
The point of data saturation	The point where no new information is expected to be discovered from further or additional investigation (e.g., further probing in an interview or additional interviews). The benefit of collecting data beyond the point of saturation is much smaller.	The number of interviews at which saturation is reached may vary depending on the study, the detail and volume of data available and the variability of the sample.[3] A general rule to follow is to "...keep asking as long as you are getting different answers..." — Harry Wolcott.[4]	When engaging ten individuals within a team, it is likely that the tenth interview provides significantly less new information than the first interview.
Expected variability in stakeholder views	The expected range of beliefs and perspectives among stakeholders may vary across initiatives.	A larger sample size should be considered if a high level of variability in stakeholder views is expected.	If the researcher expects there to be multiple viewpoints across stakeholders, a larger sample may help to capture a wider range of perspectives.

[3] Based on 60 interviews with a relatively homogenous sample, saturation of themes was reached by the twelfth interview in Guest, Bunce and Johnson (2006). However, Hagaman and Wutich (2017) found that sample sizes of 20 to 40 interviews were required to achieve data saturation of meta-themes that cut across research sites. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods. 2006;18(1):59–82; Hagaman AK, Wutich A. How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on guest, bunce, and Johnson's (2006) landmark study. Field Methods. 2017;29(1):23–41.

[4] Baker, S. E., & Edwards, R. (2012). How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research. Retrieved from http://eprints.ncrm.ac.uk/2273/.

Term	Description	Sampling considerations	Example
The stakeholders and stakeholder groups that would be useful to engage	The number of stakeholders or stakeholder groups that have been involved with, impacted by, or are interested in the initiative.	A greater sample size may be considered where there are multiple stakeholders or stakeholder groups that should be engaged.	Engage multiple groups of stakeholders, including agency executives and operational staff, initiative recipients and other stakeholders impacted by the initiative. For instance, it may be reasonable to engage at least one stakeholder from each group, or several individuals if it is a large group.

Sampling error and sampling bias

Ensuring a suitable sample composition and size helps to reduce sampling error and sampling bias. These terms are defined in Table 5.

Table 5: Sample size and composition

Term	Description	Sampling considerations	Example
Sampling error	The difference between the outcome in the sample, and the outcome in the population. As the outcome in the population is likely to be unknown, the size of the sampling error may not be known.	Sampling error will occur regardless of sampling approach, unless the sample contains all units of the population. To reduce the size of this error, increase the sample size and improve the sample composition to represent the population.	A sample may have an average score of 73. If the population average score was 75 (if measured), this would be a sampling error of two points. In practice the sampling error would not be measurable.
Sampling bias	A bias in which some members of the population have a lower or higher probability of being sampled than others. This may result in a sample that is not representative of the population (sampling error).	Where possible, the sample composition should match the population composition to reduce this bias.	In a survey, certain groups may be less likely to respond than others, which results in them being under-sampled.

Weighting

In some circumstances, it may be appropriate to weight the results obtained from a sample to better align with the initiative and its aim. This is often triggered by sample imbalances, which can result from a range of factors including communication barriers or internet access.

Weighting is commonly used to reduce sample bias and can also result in the sample better representing the entire population. Weighting, however, can inadvertently generate additional bias and increase the variability of responses. Only evidence-based weighting approaches should be used. In addition:

- for transparency purposes, all results should be reported (i.e. with and without applying weightings)
- sensitivity tests should be undertaken on subject weightings
- all modifications to the original dataset and assumptions, should be well-documented.

Rate of response

Not everybody contacted will be willing or able to respond to requests for surveys and interviews. Some members of the population may be hard to reach, due to personal characteristics, location, internet access or other communication barriers. This will impact the final sample size for which data is collected.

Low response rates are common among certain cohorts. It is important to plan for this from the start. Approaches to improve response rates may include:

- sharing the invitation/survey from an internal contact (as opposed to an external organisation)
- personalising the invitation/survey, making it easy to read and making it clearly relevant to the potential participant/respondent
- summarising what the meeting/survey will involve and the reason for the participants' involvement
- making the survey short and easy to complete (for instance, this may involve editing for plain English), or keeping the duration of the meeting to a minimum
- conducting tests with a focus group or pilot to ensure the engagement can be completed within relevant timeframes and questions are easily and correctly interpreted
- ensuring language and approach is culturally appropriate (see *Workbook IV. Evaluation plan: Manage the evaluation*)
- ensuring confidentiality and dignity of respondents when collecting their data and analysing and reporting findings (see *Workbook IV. Evaluation plan: Manage the evaluation*)
- being aware of, and supporting the participation of, cohorts that are likely to have a lower response rate (e.g., youth)
- providing incentives or rewards for participation
- emphasising the benefits of a high response rate (e.g., improved evidence to inform decisions)
- sending follow ups or reminders
- ensuring the right medium for the cohort (e.g., ensuring a cohort has internet access before sending an online survey)
- making reasonable adjustments to facilitate respondents' involvement
- distributing the invitation/survey at suitable times (e.g., quieter periods but not during holidays).